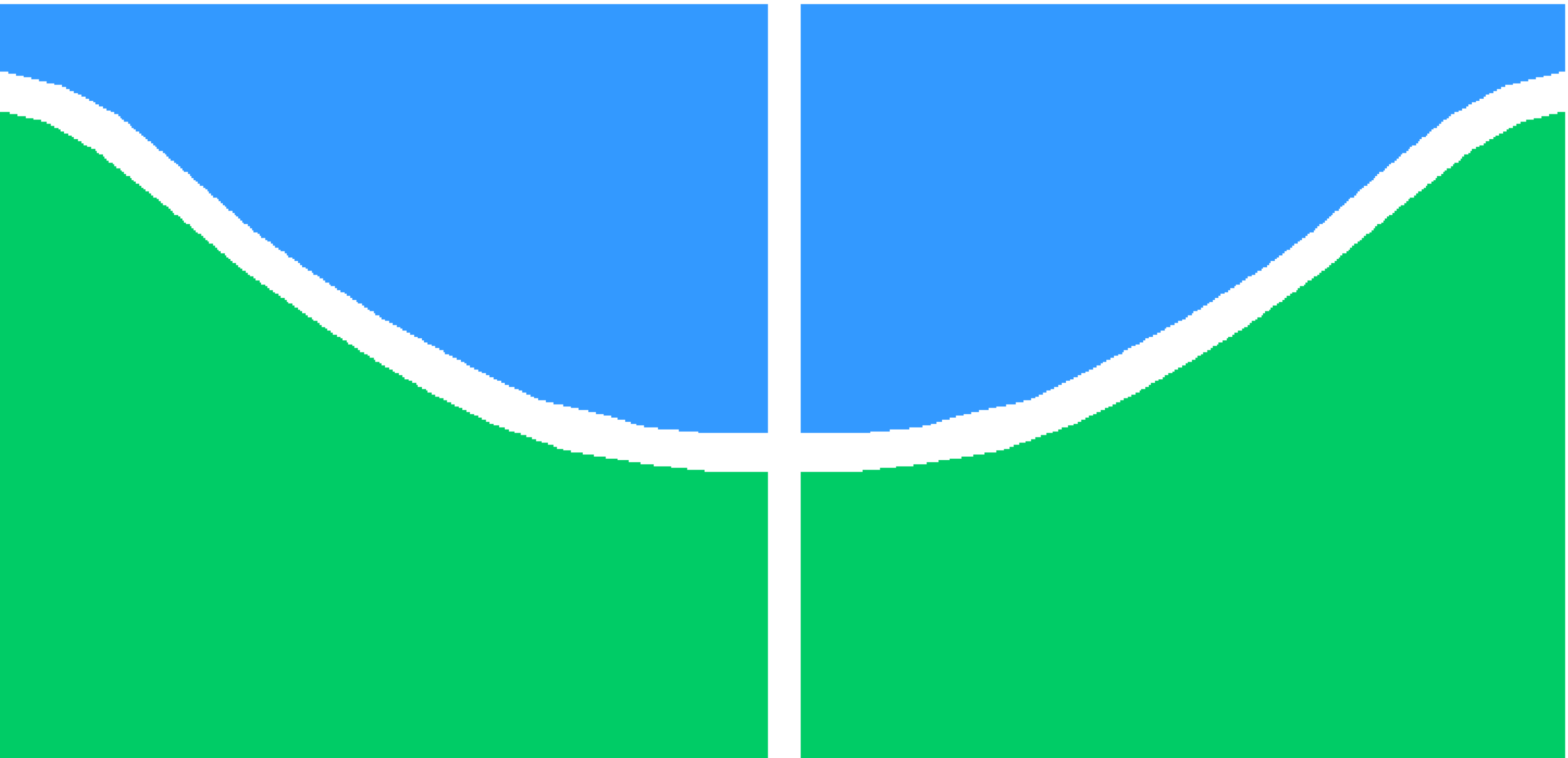


Universidade de Brasília - UnB
Faculdade UnB Gama - FGA
Engenharia de Sofware

Análise de sentimentos do *Twitter* como suporte aditivo para a previsão da volatilidade do *Bitcoin*

Autor: Nicacio Arruda Bezerra Neto
Orientador: Prof. Dr. Glauco Vitor Pedrosa

Brasília, DF
2018



Nicacio Arruda Bezerra Neto

Análise de sentimentos do *Twitter* como suporte aditivo para a previsão da volatilidade do *Bitcoin*

Monografia submetida ao curso de graduação em (Engenharia de Sofware) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia de Sofware).

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Prof. Dr. Glauco Vitor Pedrosa

Brasília, DF

2018

Nicacio Arruda Bezerra Neto

Análise de sentimentos do *Twitter* como suporte aditivo para a previsão da volatilidade do *Bitcoin*/ Nicacio Arruda Bezerra Neto. – Brasília, DF, 2018-

60 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Glauco Vitor Pedrosa

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB
Faculdade UnB Gama - FGA , 2018.

1. Palavra-chave01. 2. Palavra-chave02. I. Prof. Dr. Glauco Vitor Pedrosa. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Análise de sentimentos do *Twitter* como suporte aditivo para a previsão da volatilidade do *Bitcoin*

CDU 02:141:005.6

Nicacio Arruda Bezerra Neto

Análise de sentimentos do *Twitter* como suporte aditivo para a previsão da volatilidade do *Bitcoin*

Monografia submetida ao curso de graduação em (Engenharia de Sofware) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia de Sofware).

Trabalho aprovado. Brasília, DF, 06 de dezembro de 2018:

Prof. Dr. Glauco Vitor Pedrosa
Orientador

Prof. Dr. Fernando William Cruz
Convidado 1

Prof. Dr. Wander Cleber Maria Pereira da Silva
Convidado 2

Brasília, DF
2018

Agradecimentos

Primeiramente, agradeço a minha mãe, Suely, por ter me apoiado durante o curso de graduação e ter me dado forças para continuar estudando. Agradeço a minha esposa, Kadrynna e filhos, Arthur e Davi, pela paciência, por serem compreensíveis nos momentos de minha ausência e pelo constante apoio e motivação. Por fim, agradeço ao professor orientador Glauco Vitor Pedrosa, pelos esclarecimentos prestados durante a concepção deste trabalho.

Nicácio Arruda Bezerra Neto

Resumo

As redes sociais se tornaram fontes importantes para extração e monitoramento de informações essenciais para a gestão e o planejamento de estratégias. Para o mercado financeiro, por exemplo, a possibilidade de acompanhar as repercussões diárias sobre um ativo ou investimento é uma fonte valiosa para auxiliar os investidores em suas tomadas de decisões. Contudo, é humanamente inviável acompanhar todas as notícias sobre uma determinada aplicação financeira, por isso, um sistema automatizado que possa analisar as notícias do mercado, avaliando o sentimento em tempo real é uma ferramenta valiosa e potencialmente lucrativa. Este trabalho apresenta o desenvolvimento de uma ferramenta para analisar e exibir o sentimento das publicações relacionadas à moeda *Bitcoin* na plataforma *Twitter*. Essa ferramenta está disponível de forma online, para auxiliar investidores e pessoas que desejam visualizar a volatilidade do *Bitcoin* juntamente com sentimentos dos usuários do *Twitter* sobre esta criptomoeda.

Palavras-chaves: mineração de dados, análise de sentimentos, redes sociais, criptomoedas, twitter, bitcoin.

Abstract

Social networks have become important sources for extracting and monitoring essential information for strategy management and planning. For the financial market, for example, the possibility of tracking the daily repercussions on a financial asset or investment is a valuable source to assist investors in their decision-making. However, it is humanly impossible to keep track of all the news about a given financial application, so an automated system that can analyze the market news, evaluating real-time sentiment is a valuable and potentially profitable tool. This work aims to develop a tool that can analyze and display the sentiment of publications related to *Bitcoin* currency on *Twitter*, and also publish this information online, to assist investors and people who wish to trade *Bitcoin*. At the end of this paper, it was confirmed the relationship between the variation of the price of the currency and the publications on *Twitter*, but it was not possible to prove that the variation is caused by the feeling present in the publications, which reinforces the understanding that the sentiment expressed by the users consists largely of a reflection of the current market.

Key-words: data mining, sentiment analysis, social media, cripto-currency, twitter, bitcoin.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

API	Interface de Programação de Aplicativos
BTC	Bitcoin
BlockChain	Cadeia de blocos
Daemon	Sistema que não depende de interação com o usuário, é executado pelo sistema operacional em background
Deploy	Implantação de um software
Exchange	Plataforma de cambio
JSON	JavaScript Object Notation
MicroBlogging	Tipo de blog onde as pessoas postam mensagens mais curtas do que em blogs convencionais
OHLC	Open, High, Low e Close. Formato de dados que representa os preços de abertura, máximo, mínimo e fechamento em um determinado período.
REST	Representational State Transfer, ou Transferência de Estado Representacional
Spread	Diferença entre os preços de compra e venda de um ativo financeiro
Streaming	Forma de distribuição contínua de dados
Tweet	Nome dado a uma publicação na rede social Twitter
VADER	Valence Aware Dictionary and sEntiment Reasoner

Sumário

1 Introdução

1.1 Contextualização

A recente revolução tecnológica com a popularização das redes sociais criou uma situação sem precedentes no contexto de geração de dados, mudando drasticamente a forma de entender e analisar as ciências sociais e econômicas. A interação de pessoas com redes sociais está gerando grandes conjuntos de dados que documentam as opiniões e comportamentos de seus usuários. A exploração desses dados possibilita o descobrimento de interesses, preocupações e intenções das pessoas em relação a vários fenômenos econômicos, políticos e culturais (RANCO et al., 2015).

Há um movimento crescente na indústria em conduzir análises e fazer previsões com base em dados de mídia social. Com o acúmulo de dados e desenvolvimento de novas ferramentas para analisar e conectar grandes conjuntos de dados, técnicas de *Big Data*, mineração de dados, análise preditiva e aprendizado de máquina, estão sendo utilizadas na tentativa de entender a relação entre o comportamento humano e as tendências do mercado financeiro. (KARPPI; CRAWFORD, 2016)

Dentre todas as redes sociais existentes, o *Twitter* é hoje uma das plataformas de *microblogging* mais utilizadas. Por meio do *Twitter*, é possível verificar em tempo real o que as pessoas estão discutindo sobre determinado assunto, o volume de pessoas comentando sobre algo e as repercussões das principais notícias. Com este tipo de dado é possível explorar o comportamento, intenções e sentimentos dos usuários. A possibilidade de coletar e analisar o comportamento das pessoas é de grande interesse para várias áreas.

Para o mercado de investimentos, por exemplo, a possibilidade de verificar o relacionamento entre a opinião comum das pessoas e a movimentação, volume de negociação e demais dados financeiros sobre determinado ativo é uma fonte valiosa para o suporte na tomada de decisão. Empresas financeiras e pesquisadores acadêmicos começaram recentemente a estudar o valor preditivo das informações coletadas nas mídias sociais (KRISTOUFEK, 2013) (BUKOVINA; MARTICEK, 2016).

1.2 Justificativa

A correlação de dados de mercado com os dados de redes sociais, traz a possibilidade para formar suposições, criar previsões e agir de acordo com a previsão. Mas também é um exemplo da atual tendência de tentar modelar e compreender os desejos das pessoas e o significado de eventos no mundo através da mineração de dados de mídia social.

Há um grande interesse na construção de ferramentas de análise e apoio, especialmente quando os resultados esperados podem render lucros significativos para os portfólios dos investidores.

Existe uma discussão na literatura especializada em que pesquisadores, investidores e pessoas ligadas ao *Bitcoin* investigam a relação entre as notícias e publicações em portais de notícias e redes sociais, e a variação do preço da moeda. A partir de uma análise do material publicado que relaciona análise de sentimento e o *Bitcoin*, percebeu-se que, grande parte das publicações analisadas mostram que existe uma correlação entre o sentimento presente nas redes sociais e o comportamento deste ativo.

Partindo-se do princípio de que existe uma relação entre o sentimento presente em publicações de redes sociais e o preço do *Bitcoin*, uma ferramenta que pudesse analisar o sentimento e exibir essas informações seria de grande valor para investidores e pessoas que desejam negociar esse ativo. Essa ferramenta serviria de apoio para previsão do preço do moeda, bem como especulação do seu valor. Esse trabalho visa a fornecer indicadores de análise de sentimento presentes rede social *Twitter* em tempo real.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo desse trabalho foi desenvolver uma ferramenta computacional que pudesse realizar, de forma automática, a mineração de opiniões (análise de sentimentos) sobre a cripto-moeda *Bitcoin* usando o *Twitter* como fonte de dados. A proposta é que esta ferramenta pudesse exibir estas informações em tempo real por meio de uma pagina *web*.

1.3.2 Objetivos Específicos

Os objetivos específicos desse trabalho foram:

- Desenvolver uma ferramenta que possa coletar publicações relacionadas a moeda *Bitcoin* no *Twitter*;
- Realizar de forma automática a análise de sentimentos sobre cada *tweet* coletado, classificando-os em comentários positivos, negativos ou neutros;
- Criar uma interface gráfica que exiba a variação do volume de comentários positivos ao longo do tempo sobre o *Bitcoin*;
- Exibir as informações acerca do preço da moeda, juntamente com as informações de sentimento medidas;

- Exibir os resultados em tempo real por meio de uma página web.

1.4 Organização do Trabalho

O texto desse trabalho está organizado da seguinte forma:

- Capítulo 2: apresenta os conceitos e definições necessários para o entendimento do trabalho, bem como, uma análise dos trabalhos correlatos.
- Capítulo 3: apresenta uma descrição técnica da implementação do software, descrevendo os procedimentos e detalhes de construção da ferramenta.
- Capítulo 4: mostra os resultados obtidos através do teste experimental realizado com dados reais extraídos do *Twitter*;
- Capítulo 5: finaliza com a conclusão, algumas discussões sobre o trabalho desenvolvido e sugestões para trabalhos futuros.

2 Conceitos e Revisão Bibliográfica

2.1 Bitcoin

O *Bitcoin* é uma criptomoeda lançada em 2009. Consiste em uma moeda digital que pode ser transferida de usuário para usuário sem a necessidade de intermediários, pois possui uma rede descentralizada que processa e registra as transações, chamada *Blockchain* (Cadeia de Blocos).

As transações realizadas com essa moeda digital são verificadas através de criptografia pelos nós da *Blockchain* e registrados de forma que não seja possível alterar dados que já foram armazenados. A *Blockchain* é formada por vários usuários, também chamados de nós, espalhados pela rede, e não possui uma entidade central ou controladora.

Essa moeda digital possui código aberto. Ela é obtida através de um processo conhecido como mineração, onde, mineradores utilizam grande poder computacional para manter os registros da *Blockchain* atualizados. O *Bitcoin* tem sido utilizado como investimento, pois, seu preço sofre grande influência sob os movimentos de especulação, oferta e demanda nas *Exchanges* (casas de câmbio), gerando uma grande volatilidade.

2.2 O Twitter

O Twitter é uma rede social no formato de microblog criada em 2006. A plataforma incorpora aspectos de outras redes sociais, como Myspace e Facebook, com tecnologias de mensagens instantâneas no intuito de criar redes de usuários, os quais postam e interagem com mensagens conhecidas como *tweets*. Cada *tweet* possui um tamanho máximo de 280 caracteres.

O Twitter se tornou um novo meio de compartilhamento de informações. Segundo (KWAK et al., 2010), A grande maioria dos *tweets* publicados (mais de 85%) estão relacionados a manchetes ou notícias. *Hashtags* são usados em *tweets* para categorizar e também encontrar informação. Elas possibilitam que os usuários criem comunidades de pessoas interessadas no mesmo tópico, tornando mais fácil para que eles encontrem e compartilhem informações relacionadas.

Operadores financeiros acompanham o Twitter e o usam para prever movimentos nos mercados financeiros, pois essa plataforma pode fornecer informações úteis sobre a atividade iminente do mercado financeiro em tempo real. (KARPPI; CRAWFORD, 2016). Através dessa plataforma, notícias que impactam o mundo financeiro são postadas, compartilhadas e discutidas entre os usuários.

Em relação às mídias sociais, o Twitter está se tornando uma plataforma de microblog cada vez mais popular na análise e previsão financeira. Existem diversas pesquisas que utilizam o Twitter como fonte de dados. Elas se dividem nas seguintes linhas de pesquisa:

- Relação entre o volume de publicação e o interesse
- Relação entre indicadores de sentimento e movimentos no mercado
- Comparação entre modelos baseados no volume de publicação e modelos baseados na análise de sentimento

2.3 Mineração de dados e Aprendizado de Máquina

A mineração de dados é definida como o processo de descoberta de padrões em base de dados, onde os dados são armazenados eletronicamente e a análise e pesquisa são aceleradas ou automatizadas através de um computador. Possui uma coleção de técnicas, modelos e processos automáticos utilizados para extrair conhecimento previamente desconhecido de um grande conjunto de dados. Seu principal objetivo é criar programas que analisem os dados automaticamente, buscando regularidades e padrões. O encontro de padrões consistentes podem ser generalizados produzindo afirmações e predições.

O processo pode ser automático ou semiautomático. Os padrões obtidos podem ser significativos, e desta forma, podem trazer grandes vantagens. O aprendizado de máquina provê a base técnica para a mineração de dados.

Existem diversos tipos de mineração de dados, como a Mineração de Textos, Mineração Web, Mineração de Imagens, Mineração de Mídia Social, Mineração de Dados Espaciais, Mineração de Dados Ambientais, Mineração na Área da Saúde e Mineração em Pesquisas Genéticas. Este trabalho irá utilizar a **Mineração de Textos**, **Mineração Web** e **Mineração de Mídia Social**. Abaixo encontra-se uma pequena descrição dessas modalidades de mineração de dados.

2.3.1 Mineração de Textos

Processo de obtenção de informações de alta qualidade a partir de textos. Informações de alta qualidade podem ser obtidas a partir do aprendizado de padrões estatísticos. A mineração de texto geralmente envolve o processo de estruturação do texto de entrada (analisando e adicionando algumas características linguísticas derivadas, bem como, a inserção em um banco de dados), derivando padrões dentro dos dados estruturados, e finalmente avaliando e interpretando a saída. A "alta qualidade"na mineração de texto geralmente se refere a alguma combinação de relevância. As tarefas típicas de mineração

de texto incluem categorização de texto, agrupamento de texto, extração de conceito / entidade, análise de aspectos léxicos e gramaticais, produção de taxonomias granulares, análise de sentimentos, resumo de documentos e modelagem de relações entre entidades,

As primeiras aplicações de mineração de textos podiam extrair informações de documentos estruturados como documentos do Microsoft Word e Acrobat PDF. Novos desenvolvimentos em mineração de texto agora digitalizam e analisam texto não estruturado como memorandos, pesquisas, chats, notas, Fóruns e apresentações (BEGUM; BEGUM, 2013). A mineração de textos está em crescimento, áreas de interesse como a mineração de dados multilíngue está em ascensão. Nesse sentido, tem-se o objetivo de obter dados de diferentes idiomas e agrupar itens similares de diferentes fontes linguísticas de acordo com seu significado.

O grande obstáculo para a mineração de texto está em sua própria natureza, pois a forma em que é estruturado é imprópria para o processamento automatizado. (HEARST, 1999).

2.3.2 Mineração Web

Aplicação das técnicas de mineração de dados para descobrir padrões de uso de na internet, com o objetivo de entender o comportamento dos usuários e criar aplicações que melhor atenda seus interesses. Pode avaliar conteúdos da internet, através de processamento e classificação, relacionamento entre as páginas através de links de redirecionamento, e comportamentos de usuários, através de dados capturados, registro de operações, comportamentos de pesquisa.(COOLEY; MOBASHER; SRIVASTAVA, 1997).

A mineração de conteúdos da web possui grande complexidade. Isso se deve a falta de uma estrutura padronizada de postagem de conteúdos.

2.3.3 Mineração de Mídia Social

Consiste no processo de representação, análise e extração de padrões e tendências a partir de dados de mídia social, como por exemplo: Taxa de uso de mídias sociais, comportamentos online, compartilhamento de conteúdo, conexões entre indivíduos, comportamento de compra online.(ZAFARANI; ABBASI; LIU, 2014)

Algumas áreas de pesquisa:

- Detecção de eventos em redes sociais;
- Estrutura da comunidade (Detecção / Evolução / Avaliação da comunidade);
- Cascata de informações;
- Influência e homofilia;

- Recomendação nas redes sociais;
- Pesquisa social.

A mineração de mídia social enfrenta grandes desafios como a grande quantidade de dados, obtenção de amostras suficientes, falha de remoção de ruídos e alta complexidade das relações entre entidades.

2.3.4 Análise de Sentimentos

Análise de sentimento, também conhecida como *Mineração de Opinião* é um campo de pesquisa que estuda e analisa opiniões, sentimentos, avaliações, atitudes, emoções das pessoas. (LIU, 2012). Seu objetivo é entender a opinião de determinado agente e classificar essa opinião sob a perspectiva de expressão de sentimento. Para fins de exemplo, um sentimento pode ser positivo, negativo e neutro.

O crescimento da web e o surgimento de sites de publicação de opiniões tais como blog pessoais e redes sociais em geral como o Twitter, trouxeram consigo novas oportunidades para a exploração e descoberta da opinião das pessoas. Milhões de pessoas expressam suas suas opiniões por meio de redes sociais. Pela primeira vez na história da humanidade, temos um enorme volume de dados opinativos gravados em formato digital. A ańalise de sentimento ou mineração de opinião, é considerada uma das áreas de pesquisas mais ativas em processamento de linguagem natural. É amplamente estudada em mineração de dados, mineração da web e mineração de textos. (LIU, 2012)

A publicação da opinião detém de alto poder de influência sob demais indivíduos. A rede social pode influenciar o comportamento e a tomada de decisão das pessoas.

2.3.4.1 Ferramentas para Análise de Sentimentos

Em análise de sentimento, inicialmente deve-se definir um conjunto de palavras que expressam os sentimentos. Esse conjunto de palavras é denominado léxico de sentimentos; Definir esse conjunto de palavras é demasiadamente trabalhoso, de modo que, já existem diversos léxicos de sentimentos que foram construídos. Exemplo:

- Léxico de Biu Liu - Opinion Lexicon
- MPQA Subjectivity Lexicon
- SentiWordNet
- Harvard General Inquirer
- LIWC

Existem dois tipos principais de Léxicos de sentimentos: Os que se baseiam na polaridade, e os que se baseiam na valência. Nos baseados em polaridade, os textos são avaliados em positivos e negativos. Nos Léxicos baseados em valência, além da avaliação da polaridade, a intensidade do sentimento também é levada em consideração. Por exemplo, as palavras "bom"e "excelente"seriam tratadas da mesma forma em uma abordagem baseada na polaridade, enquanto que, na abordagem baseada em valência, a palavra "excelente"seria tratada como mais positiva do que a palavra "bom".

A análise de textos publicados e redes sociais possui alguns desafios e particularidades:

- Frequência e volume de dados gerados
- Uso de abreviações e convenções de linguagem
- Redução de contexto

Os léxicos listados acima possuem diferentes características. Um novo léxico proposto em (HUTTO; GILBERT, 2014), foi construído com foco no domínio de redes sociais do tipo microblog, como o Twitter.

O léxico VADER (for Valence Aware Dictionary for Sentiment Reasoning) foi construída com foco em análise de sentimento em plataformas de microblog como o twitter, abaixo estão algumas características desse léxico:

- Focado em microblog;
- Classifica gírias, expressões de linguagem, abreviações e até emoticons
- Combinação qualitativo e quantitativo;
- Atribui uma pontuação de valência para intensidade do sentimento
- Uso validado no twitter com 96% de precisão;
- Testes revelaram maior precisão do que avaliadores humanos;

Ao analisar um *tweet* ou sentença com VADER, temos um retorno abaixo: 'pos': 0,0, 'compound': -0,5461, 'neu': 0,221, 'neg': 0,777, onde *pos*, *neu* e *neg* representam a proporção do texto classificada como sentimento positivo, neutro e negativo respectivamente.

O atributo compound é calculado somando todos os escores de valência de cada palavra no léxico. Este atributo varia entre -1 (extremamente negativo) e 1 (extremamente positivo).

Através desse atributo pode-se estabelecer limites padronizados para a classificação de sentimento. Por exemplo: sentimento positivo : compound >= 0,05 sentimento neutro : (compound > -0,05) e (compound <0,05) sentimento negativo : compound <= -0,05. Na tabela 2.3.4.1 é possível observar alguns testes realizados com o Vader.

Sentença	% Negativa	% Neutra	% Positiva	Sentimento Composto
Bitcoin is a great alternative for international transfers	0.0	0.594	0.406	0.6249
Bitcoin sucks, I lost my money.	0.615	0.385	0.0	-0.5859
Bitcoin has a great future	0.0	0.423	0.577	0.6249
It's a bad idea to sell bitcoin now.	0.368	0.632	0.0	-0.5423
Transfers in Bitcoin are fast and cheap	0.0	1.0	0.0	0.0

Tabela 1 – Testes de sentenças com Vader

O $Vader$ analisa cada palavra na sentença e verifica a existência dessa palavra no Léxico. Caso exista palavras na sentença que existam no Léxico, ele atribui o valor de sentimento para esta palavra. Pode-se notar que grande parte as sentenças (4) foram classificadas corretamente, porém, sabemos que, para um analisador humano, a última sentença seria classificada como positiva. Porém, o $Vader$ não encontrou nenhuma palavra que esteja presente em seu Léxico, o que, acabou tornando o julgamento da sentença como neutra;

2.4 Análises Técnicas de Tendências de Mercado

Os investidores experientes costumam medir as tendências do mercado para então poder tirar proveito delas como suporte à tomada de decisão. A seguir é apresentada uma técnica, comumente utilizada por investidores, para inferir uma tendencia no comportamento da volatilidade no preço de um determinado ativo financeiro.

2.4.1 Candlestick - Gráfico de Vela

A análise de *candle* foi inicialmente desenvolvida por Munehisa Homma no início do século XVIII, para a análise de contratos futuros de arroz, no primeiro mercado contratos futuros, a Dojima Rice Exchange em Osaka no Japão. A introdução da análise de *candle* no mundo ocidental se deu em 1991 por (NISON, 1991). Em seu artigo *Japanese Candlestick Charting Techniques*, Nilson notou que poderia usar o gráfico em outras aplicações financeiras como moedas, ações e contratos futuros. No Brasil, o gráfico de *candle* é conhecido como gráfico de velas.

Os gráficos de vela utilizam quatro indicadores técnicos em sua construção. São eles:

- Preço de Abertura: é o primeiro preço inicialmente divulgado do ativo, durante o período da vela.

- Maior Preço: se refere ao maior preço que o ativo atingiu durante o período da vela.
- Menor preço: é o menor preço que o ativo atingiu durante o período da vela.
- Preço de Fechamento: é o preço final, o último preço divulgado do ativo, durante o período da vela.

Um exemplo de gráfico de vela está ilustrado na figura 1. Cada vela possui três elementos principais, são eles:

- Período: a vela representa as variações no preço (abertura, maior preço, menor preço e fechamento) durante um período de tempo determinado. Este período pode assumir vários valores. Exemplo: 1 hora, 1 minuto, 5 minutos , 1 mês.
- Formato: o formato é determinado da seguinte maneira: O corpo é representado pelos preços de abertura e fechamento. As linhas verticais ligadas ao corpo representam os preços mínimo e máximo.
- Cor: a cor da vela determina se o preço do ativo aumentou ou diminuiu. Quando o preço de fechamento é maior do que o preço de abertura, significa que durante o período de tempo da vela o preço aumentou. Se o preço de fechamento for menor do que o de abertura, significa que o preço diminuiu. Se os preços de fechamento e abertura forem iguais, então o preço final é o mesmo do início da operação.

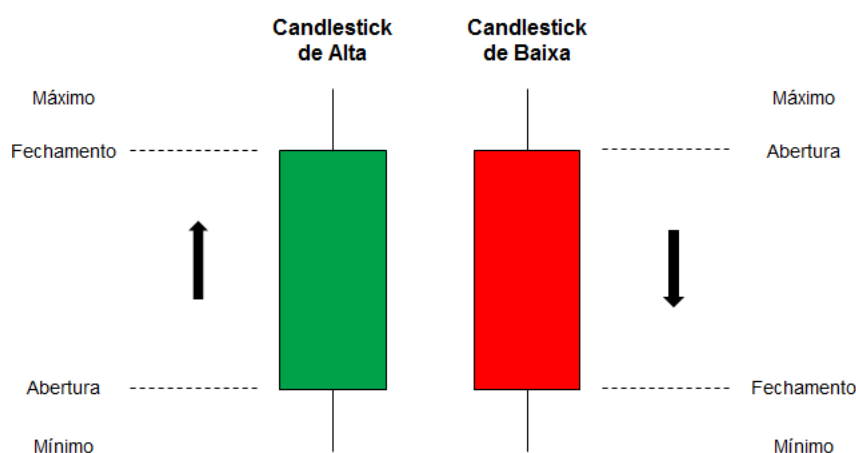


Figura 1 – Exemplo de um Gráfico de Velas

A figura 2 mostra um exemplo de gráfico de velas com valores reais do *Bitcoin* ao longo dos anos de 2016 a 2018. Como se pode visualizar, cada vela no gráfico possui um determinado tamanho, que em suma, representa a variação do preço do ativo num determinado período de tempo. Quando a vela tem um comprimento (maior preço - menor

preço) pequeno, houve pouca variação no preço durante o período da vela. Se a vela tem um comprimento maior, houve uma maior variação do preço durante o período da vela.



Figura 2 – Gráfico de Velas do BTC

Os gráficos de vela refletem não apenas a mudança entre o equilíbrio entre oferta e procura (CAGINALP; LAURENT, 1998), mas também aspectos do sentimento e psicologia do investidor (MARSHALL; YOUNG; CAHAN, 2008).

2.5 Trabalhos Correlatos

Na literatura pode-se encontrar trabalhos relacionando o comportamento dos usuários em redes sociais e a movimentação no mercado da cripto-moeda *Bitcoin*. A seguir são apresentados e discutidos alguns desses trabalhos em ordem cronológica a partir do ano de 2013.

- O trabalho de (KRISTOUFEK, 2013) mostrou que existe uma correlação entre volume de pesquisa no Google, visitas a Wikipedia e o preço do *Bitcoin* em uma deter-

minada *exchange*. Os autores identificaram uma forte correlação entre o volume de pesquisa e o preço do ativo, evidenciando um aumento do interesse ou desinteresse do público quando a moeda está acima ou abaixo de seu valor de tendência.

- (GARCIA et al., 2014) realizaram um estudo para verificar se a opinião e o sentimento das pessoas contribuíam para o surgimento de uma bolha no preço do *Bitcoin*. Para isso, foram verificados os traços comportamentais coletivos e fenômenos sociais nas plataformas Twitter, Wikipédia e Facebook. Analisando o compartilhamento de informações de usuário para usuário no Twitter, Facebook e pesquisas na Wikipédia, (GARCIA et al., 2014) encontraram dois ciclos de feedback, um ciclo social e um ciclo de adoção do usuário. No ciclo social, usuários trocam informações entre si e no ciclo de adoção, novos usuários começam a adotar a moeda. Como resultado dessa pesquisa, (GARCIA et al., 2014) observaram que picos na busca de informações, presumivelmente ligados a eventos externos, precediam reduções drásticas de preços da moeda.

- (KAMINSKI; GLOOR, 2014) analisaram o sentimento presente em conjunto de *tweets* durante 104 dias. Para análise de sentimento, os autores utilizaram um dicionário com pequeno número de palavras (15) para representar sentimentos positivos e negativos. Após isso, aplicou correlações de Pearson para buscar possíveis correlações entre o preço e volume diário da moeda. (KAMINSKI; GLOOR, 2014) descobriram que os *tweets* negativos e os *tweets* que possuíam sentimento de incerteza se relacionavam de modo proporcional ao volume de negociação e inversamente proporcional ao preço. A inserção de um atraso mostrou que altos volumes de negociação precedem aumento de *tweets* negativos e com sentimento de incerteza; ,Após a descoberta dessas correlações, foi aplicado a causalidade de Granger para verificar a confirmação das correlações. Entretanto, a aplicação desta técnica não confirmou as correlações descobertas.

- (GEORGOULA et al., 2015) usa análise de séries temporais para estudar a relação entre preços de *Bitcoin* juntamente com análise de sentimento no Twitter e pesquisas no Wikipedia. Na análise em questão o autor utilizou Máquina de Vetores de Suporte (SVMs) como classificador, onde uma série de regressões mostraram que o sentimento e o grau público de interesse esteve positivamente correlacionado com os preços da moeda.

- (GARCÍA; SCHWEITZER, 2015) revelaram que as elevações na análise de sentimento precedem aumentos na polarização de opinião e no volume de negociação. Por sua vez, um aumento desses fatores precede um aumento nos preços do *Bitcoin*.

- (MATTA; LUNESU; MARCHESI, 2015), examinou se o *spread* (diferença entre o preço de compra e venda) do *Bitcoin* tem algum relacionamento com os volumes de

tweets ou resultados de pesquisas na Web. Os autores compararam as tendências de preço e dados do Google Trends, volume de *tweets* e *tweets* com sentimento positivo.

- (BUKOVINA; MARTICEK, 2016) produziram um estudo sobre os fatores que influenciam a volatilidade excessiva do *Bitcoin*. Além disso, os autores propuseram uma decomposição do preço da moeda em fatores puramente racionais e menos racionais. Através de análise de sentimento, mostraram que o sentimento positivo é mais influente para a volatilidade excessiva do *Bitcoin*.

- O trabalho de (KARALEVICIUS; DEGRANDE; WEERDT, 2018), buscou medir a interação entre o sentimento expresso na mídia diariamente e o preço do *Bitcoin*. Para isso, utilizaram um banco de dados de artigos e notícias relativas ao tema, bem como postagens em blogs. A análise de sentimento foi baseada no Harvard Psychosocial Dictionary e um dicionário voltado para finanças. Esse trabalho propôs uma estratégia de *trading* com base no sentimento medido durante o dia. A estratégia proposta produziu resultados relativamente positivos. (KARALEVICIUS; DEGRANDE; WEERDT, 2018) constataram a existência de uma relação entre o preço e o sentimento medido, onde, após a publicação de uma notícia impactante, o primeiro preço vai na direção do sentimento expresso na notícia, porém o mercado tende a reagir com um movimento de correção.

3 Desenvolvimento

3.1 Metodologia de desenvolvimento

A metodologia utilizada para o desenvolvimento do projeto foi o *Kanban*, que é uma metodologia ágil simplificada, flexível e que possui uma grande vantagem: a visualização do fluxo de trabalho. Essa metodologia adequou-se bem ao projeto, pois, é uma metodologia de rápida implantação, e que pode ser utilizada em projetos que possuem apenas um desenvolvedor.

Um dos fundamentos da metodologia *KanBan* é a criação de um modelo visual do fluxo de trabalho (um quadro de tarefas). Esse modelo visual não possui uma estrutura fixa, porém, geralmente busca-se mostrar as tarefas que serão realizadas, que estão sendo realizadas e que já foram realizadas. Dessa forma, a percepção do progresso das tarefas do projeto é facilmente compreendida, bastando uma rápida visualização desse quadro. A partir desse modelo visual também é possível fazer a priorização das tarefas, uma vez que, a própria organização das tarefas no quadro indicam sua prioridade.

A limitação do trabalho em andamento é outro fundamento importante dessa metodologia. Através da divisão do quadro em faixas como (a fazer, em progresso e finalizadas), é possível limitar a quantidade de tarefas presente em cada faixa, controlando o fluxo de trabalho.

3.2 Requisitos e Arquitetura da Ferramenta

Na tabela 2 encontra-se uma descrição dos requisitos da ferramenta desenvolvida em alto nível de abstração.

A ferramenta deve realizar a coleta das informações de forma ininterrupta. En-

Identificador	**Nome**	**Descrição**
RF01	Coletar Tweets	O sistema deve ser capaz de receber as publicações sobre o Bitcoin no Twitter em tempo real. O sistema deve estar preparado para receber os dados a qualquer momento.
RF02	Armazenar Tweets	O sistema deve ser capaz de armazenar a data de criação e o texto de publicação de cada tweet, descartando os dados que não são necessários.
RF03	Análise de Sentimentos	O sistema deve ser capaz de realizar análise de sentimento de cada tweet armazenado. A análise de sentimento deve ser realizada a cada novo período de uma hora.
RF04	Gráfico de Velas	O sistema deve ser capaz de exibir o gráfico de velas do Bitcoin em tempo real, utilizando como período de vela uma hora.
RF05	Indicador de Sentimentos	O sistema deve exibir um indicador de sentimento em tempo real, utilizando como tempo mínimo, o sentimento presente em um período de uma hora.
RF06	Exibição dos Resultados	Os resultados do gráfico de velas e da análise de sentimento devem ser disponibilizados por meio de um website, de modo que qualquer usuário conectado a internet possa ter acesso

Tabela 2 – Requisitos do Sistema Desenvolvido

quanto que, um subprocesso inicializado por uma CRON (mecanismo de agendamento de tarefas) terá de realizar a análise de sentimento das publicações do período, coletar os dados financeiros do *Bitcoin* (valor inicial, valor final, valor máximo e valor mínimo) no período, construir os gráficos e finalmente, realizar a publicação dos gráficos online. O subprocesso deve ser executado em loop, de modo que, ele deve ser inicializado ao fim de cada período de uma hora. Exemplo:

A coleta de publicações iniciou as 00:00 do dia 01/08, às 01:00, o primeiro período se encerrou, então, exatamente após o término do período, a CRON inicializou o subprocesso, dando inicio a análise de sentimento das publicações deste período que se encerrou. Ao fim do subprocesso, há um tempo de espera para que se termine o próximo período, ou seja, às 02:00, a CRON novamente inicializa o subprocesso.

A partir da análise dos requisitos e do processo, decidiu-se que a ferramenta seria composta por três subsistemas:

- Subsistema 1 - Coletor de Tweets
- Subsistema 2 - Analisador de dados
- Subsistema 3 - Aplicação Web

em que, o subsistema 1 irá abranger os requisitos RF01 e RF02, o subsistema 2 irá abranger os requisitos RF03, RF04 e RF05, e o subsistema 3 irá abranger os requisitos RF04, RF05 e RF06. Os dados utilizados neste trabalho se referem aos *tweets* e aos valores dos preços do *Bitcoin*. A seguir são descritos os detalhes técnicos utilizados na implementação de cada subsistema que compõe a ferramenta.

3.3 Subsistema 1 - Coletor de Tweets

3.3.1 Responsabilidades

Este subsistema tem a responsabilidade de coletar as publicações, realizar o pré-processamento e armazenar os dados pré-processados no banco de dados.

3.3.2 Coleta de Tweets

Inicialmente, o subsistema se comunica com *API* (Interface de Programação de Aplicativos) do Twitter, enviado as credenciais para abertura de uma conexão e as palavras chave para que a *API* realize um filtro nas publicações que serão enviadas. Ao receber cada publicação, ele realiza o pré-processamento, armazenando apenas o texto e a data de publicação.

Uma conexão *streaming* (transmissão contínua), é utilizada para receber as publicações, ou seja, os dados são enviados dos servidores do *Twitter* e recebidos por um servidor dedicado de maneira ininterrupta. Esse software é executado em modo *daemon* (executado em plano de fundo, não é necessário interferência humana).

Ao invés de realizar diversas requisições para *API* e receber os dados em lotes, como em uma *API REST* (Transferência de Estado Representacional), através da abertura de uma única conexão com a *API* de *streaming* do *Twitter*, os dados são enviados e recebidos até que a conexão seja encerrada por uma das partes. Isso permite com que os dados sejam entregues com baixa latência e alta taxa de transferência.

Os dados são enviados e recebidos em formato *JSON* (JavaScript Object Notation). Para facilitar a coleta dos dados, foi utilizado a biblioteca **twitter**. Essa biblioteca facilita a interação e autenticação com a *API* do *Twitter* para programas escritos na linguagem **Ruby**.

Para realizar a filtragem dos *tweets*, foram definidas as seguintes palavras-chaves:

- Bitcoin
- BTC
- Satoshi

Através dessas palavras-chaves, qualquer conteúdo publicado no Twitter, escrito em língua inglesa e que tivesse alguma dessas palavras, é capturado pelo coletor de *tweets*. Palavras compostas também são consideradas e não há diferenciação entre letras maiúsculas e minusculas

3.3.3 Pré-processamento dos Tweets

A Figura 3 mostra a estrutura de um *tweet*. Na estrutura do *tweet*, tem-se outras subestruturas, como as informações de user (Informações do usuário que postou o conteúdo), e extended tweet (Informações detalhadas sobre o *tweet*).

```
{ ⊟
   "created_at":"Wed Oct 10 00:00:00 +0000 2018",
   "id":1049811748851716097,
   "id_str":"1049811748851716097",
   "text":"@Salvador_Diaz\nYour stats for Bitcoin Cash 📊
\n\nRank: 4\nPrice USD: 517.79$\nTrend 1h: 0.31% ↗\nTrend 24h: -1.77% ↘\nTr… https://t.co/0ODW4eSUckL",
   "source":"<a href=\"https://twitter.com/AltcoinTicker\" rel=\"nofollow\">AltcoinTicker</a>",
   "truncated":true,
   "in_reply_to_status_id":null,
   "in_reply_to_status_id_str":null,
   "in_reply_to_user_id":18196495,
   "in_reply_to_user_id_str":"18196495",
   "in_reply_to_screen_name":"Salvador_Diaz",
   "user":{ ⊞ },
   "geo":null,
   "coordinates":null,
   "place":null,
   "contributors":null,
   "is_quote_status":false,
   "extended_tweet":{ ⊞ }
}
```


Figura 3 – Estrutura do Tweet

A Figura 4 exibe a subestrutura do *tweet* contendo as informações do usuário que realizou a postagem.


```
"user":{ ⊟
   "id":904979075604901888,
   "id_str":"904979075604901888",
   "name":"AltcoinTicker",
   "screen_name":"AltcoinTicker",
   "location":null,
   "url":null,
   "description":"I tweet out the Top 5 Coins daily and can send you infos for your favourite coin.\nIf you like me, check out my Firefox Extension: \nhttps://goo.gl/5kMa7B",
   "translator_type":"none",
   "protected":false,
   "verified":false,
   "followers_count":62,
   "friends_count":0,
   "listed_count":3,
   "favourites_count":0,
   "statuses_count":5553,
   "created_at":"Tue Sep 05 08:06:40 +0000 2017",
   "utc_offset":null,
   "time_zone":null,
   "geo_enabled":false,
   "lang":"de",
   "contributors_enabled":false,
   "is_translator":false,
   "profile_background_color":"000000",
   "profile_background_image_url":"http://abs.twimg.com/images/themes/theme1/bg.png",
   "profile_background_image_url_https":"https://abs.twimg.com/images/themes/theme1/bg.png",
   "profile_background_tile":false,
   "profile_link_color":"19CF86",
   "profile_sidebar_border_color":"000000",
   "profile_sidebar_fill_color":"000000",
   "profile_text_color":"000000",
   "profile_use_background_image":false,
   "profile_image_url":"http://pbs.twimg.com/profile_images/907734954938961920/iUvYtucd_normal.jpg",
   "profile_image_url_https":"https://pbs.twimg.com/profile_images/907734954938961920/iUvYtucd_normal.jpg",
   "profile_banner_url":"https://pbs.twimg.com/profile_banners/904979075604901888/1505255925",
   "default_profile":false,
   "default_profile_image":false,
   "following":null,
   "follow_request_sent":null,
   "notifications":null
},
```


Figura 4 – Informações do Usuário

A Figura 5 exibe a subestrutura do *tweet* (extended tweet) contendo as informações complementares da publicação.

```
{ ⊟
   "created_at":"Wed Oct 10 00:00:00 +0000 2018",
   "id":1049811748851716097,
   "id_str":"1049811748851716097",
   "text":"@Salvador_Diaz\nYour stats for Bitcoin Cash 📊
\n\nRank: 4\nPrice USD: 517.79$\nTrend 1h: 0.31% ↗\nTrend 24h: -1.77% ↘\nTr… https://t.co/0ODW4eSUckL",
   "source":"<a href=\"https://twitter.com/AltcoinTicker\" rel=\"nofollow\">AltcoinTicker</a>",
   "truncated":true,
   "in_reply_to_status_id":null,
   "in_reply_to_status_id_str":null,
   "in_reply_to_user_id":18196495,
   "in_reply_to_user_id_str":"18196495",
   "in_reply_to_screen_name":"Salvador_Diaz",
   "user":{ ⊞ },
   "geo":null,
   "coordinates":null,
   "place":null,
   "contributors":null,
   "is_quote_status":false,
   "extended_tweet":{ ⊞ }
}
```


Figura 5 – Informações complementares

Dado o conjunto de informações da figura 5, as informações necessárias para a análise são:

- Data de Criação ("created at", ou "timestamp ms")
- Texto publicado ("full text")

Logo, durante o pré-processamento dos dados, é feita a remoção dos demais dados de cada *tweet*, de modo a salvar apenas o conteúdo necessário no banco de dados. A remoção dos dados é realizada através de manipulação JSON.

3.3.4 Ferramentas Utilizadas

O Coletor de Tweets foi desenvolvido na linguagem de programação **Ruby**, versão 2.5.3. A linguagem **Ruby** foi escolhida por causa da biblioteca **twitter**. Essa biblioteca funciona como uma interface facilitando a conexão com a *API* do Twitter. O banco de dados utilizado foi o **Postgres** na versão 9.6. Este banco de dados foi escolhido por ter boa performance e aceitar o formato JSON como tipo de dado.

3.4 Subsistema 2 - Analisador de dados

3.4.1 Responsabilidades

Este software tem a responsabilidade de recuperar os dados do banco de dados, ordenar os dados e realizar análise de sentimento dos dados coletados, coletar dados

financeiros do *Bitcoin*, construir os gráficos e finalmente, publicar os gráficos online.

3.4.2 Ordenação dos Tweets

Essa é a primeira tarefa realizada pelo Analisador de dados. Inicialmente, este software recupera os *tweets* de um determinado período do banco de dados e os ordena por hora. Nesta tarefa, foi utilizado um dicionário com chave = "dia-hora", valor = *tweet*, dessa forma, o acesso aos *tweets* fica mais simples, pois é possível filtrar os *tweets* por dia/hora.

A recuperação das informações (data de criação e conteúdo) também fica mais simples, uma vez que os *tweets* estão presentes do dicionário, basta acessar o *tweet* desejado e escolher o atributo.

3.4.3 Análise de sentimento dos Tweets

Essa é a segunda tarefa realizada pelo Analisador de dados. Para classificação dos *tweets* de um determinado intervalo, é criado uma lista e anexado o conteúdo ("full text") da publicação de cada *tweet* presente neste intervalo. Essa lista é passada como parâmetro para um função que irá iterar e avaliar cada sentença, calculando o sentimento individualmente através da biblioteca **vaderSentiment**. Essa biblioteca recebe como parâmetro o texto da publicação e retorna o valor do sentimento no formato:

```
{'pos': valor, 'compound': valor, 'neu':valor, 'neg':valor}.
```

Os scores **pos**, **neg** e **neu** representam a proporção da sentença que se enquadra em cada categoria (positivo, negativo e neutro). Estas medidas são utilizadas quando necessitamos de indicadores de sentimento multidimensionais. A medida **compound** é calculada somando os escores de valência de cada palavra no léxico e normalizada para estar entre -1 (extremo negativo) e +1 (extremo positivo). A media **compound** é mais indicada para estabelecer uma classificação de sentenças com limites padronizados, definindo assim, intervalos de valores para o enquadramento de sentenças positivas, negativas e neutras.

Segundo (HUTTO; GILBERT, 2014), os valores típicos utilizados na literatura para a classificação de sentenças são:

```
sentimento positivo : compound >= 0,05
sentimento neutro : ( compound > −0,05) e ( compound <0,05)
sentimento negativo : compound <= −0,05
```

Por estes motivos, a medida **compound** foi escolhida para ser utilizada como índice de classificação. Os critérios acima foram utilizados para o enquadramento de sentenças.

O cálculo do sentimento em cada intervalo é dado pela seguinte fórmula:

$$indicator = \frac{\sum_{t=0}^{tweetcount} compound(t)}{tweetcount}$$

em que, o valor de indicador consiste na razão entre o somatório do valor de sentimento, e o número de *tweets* analisados. Deste modo, o valor de sentimento não sofrerá em momentos em que poucos *tweets* forem publicados, por exemplo, em horários com baixa taxa de *tweets* publicados.

3.4.4 Construção do gráfico de velas

Para construir o gráfico de velas, foi necessário levantar as informações de mercado de *Bitcoin* correspondentes a cada hora. Para isso, foi utilizado a **API Cryptocompare**. Essa API fornece dados de negociação de diversas cripto-moedas em 181 *exchanges*. Por meio dela é possível obter dados em tempo real e dados históricos de diversas cripto-moedas. Os dados necessários são os dados disponibilizados no formato OHLC (Open, High, Low, Close), com período de uma hora.

O endpoint utilizado foi <https://min-api.cryptocompare.com/data/histohour?fsym=BTC&tsym=USD&limit=23&toTs=timestamp>, passando como parâmetro o par BTC/USD, que representa o preço do *Bitcoin* em relação ao dólar, e a data em timestamp, utilizando o mesmo fuso horário da blockchain: UTC (Tempo Universal Coordenado). Os dados do mercado são importados em formato JSON. Dessa forma, a manipulação é simplificada. A Figura 6 mostra um exemplo de resposta da *API* com essas configurações.

```
{
    "Response": "Success",
    "Type": 100,
    "Aggregated": false,
    "Data": [
        {
            "time": 1542060000,
            "close": 6373.44,
            "high": 6374.58,
            "low": 6350.87,
            "open": 6364.77,
            "volumefrom": 1202,
            "volumeto": 7655723.02
        },
        {
            "time": 1542063600,
            "close": 6375.08,
            "high": 6380.3,
            "low": 6363.83,
            "open": 6372.95,
            "volumefrom": 730.71,
            "volumeto": 4671927.25
        },
        {
            "time": 1542067200,
            "close": 6371.02,
            "high": 6377.09,
            "low": 6360.71,
            "open": 6375.08,
            "volumefrom": 525.27,
            "volumeto": 3355223.58
        }
    ],
    "TimeTo": 1542067200,
    "TimeFrom": 1542060000,
    "FirstValueInArray": true,
    "ConversionType": {
        "type": "direct",
        "conversionSymbol": ""
    },
    "RateLimit": {},
    "HasWarning": false
}
```


Figura 6 – Dados OHLC

Para construir os gráficos, é criado um arquivo no formato ".csv"com o seguinte formato:

"t0","t1","tweets count","indicator", "open", "high", "low", "close"

em que, **t0** é a hora inicial, **t1** é a próxima hora, **tweets count** representa quantos *tweets* foram analisados no intervalo t0...t1, **indicator** representa o indicador de sentimento nesse intervalo, **high** representa o maior valor da moeda durante o período, **low** representa o menor valor da moeda durante o período, **open** representa o inicial da moeda durante o período e **close** representa o valor de fechamento da moeda durante o período. Nesse arquivo, cada linha representa um intervalo, começando de 00:00 do dia XX/XXXX até o momento atual. A cada hora, ele deve ser atualizado com uma nova linha, representando um novo intervalo. Ao fim, este arquivo é utilizado para construção de um dataframe que servirá como fonte para a elaboração de dois gráficos, um gráfico de vela representando o comportamento do *Bitcoin*, e um gráfico de barras representando a análise de sentimento no determinado período.

3.4.5 Ferramentas

Este software foi desenvolvido utilizando a linguagem **Python**, na versão 3.6. Foram utilizadas as bibliotecas **psycopg2** para integração com o banco de dados **Postgres**, **datetime** para transformação de datas no formato timestamp para o formato ISO (AAAA-MM-DD HH:MM:SS), **vaderSentiment** para realizar análise de sentimento usando léxico Vader, **requests** para consumir a **API Cryptocompare** através de requisições HTTP, **json** para realizar o parse dos dados da **API Cryptocompare**, **Plotly** para construção e publicação dos gráficos online.

3.5 Subsistema 3 - Aplicação Web

3.5.1 Responsabilidades

Esta aplicação web tem a responsabilidade de exibir os resultados de análise de sentimento e variação do preço do *Bitcoin* online, facilitando a visualização e interpretação dos resultados.

3.5.2 Arquitetura

A aplicação web foi desenvolvida seguindo o padrão arquitetural MVC (Model - View - Controller), onde temos a separação entre as camadas de modelagem de dados, interface com usuário e processamento de requisições. Este padrão foi escolhido por ser escalável e por ser de fácil evolução, tendo em vista que, tem-se o objetivo de adicionar a análise de outras moedas e ativos financeiros.

Model: Camada responsável pela manipulação dos dados

View: Camada que possui interação com usuário, responsável pelas interfaces que exibem os dados da model.

Controller: Faz ligação entre as outras duas camadas, model e view, recebendo as requisições do usuário, buscando e alterando dados na model e devolvendo a view para o usuário.

3.5.3 Organização e Exibição dos Resultados

A grande maioria das plataformas de análise financeira representam seus ativos utilizando dois gráficos em uma única visualização: um gráfico de *vela*, representando a variação do valor do ativo financeiro, e abaixo tem-se um gráfico de barras, representando o volume de negociação do ativo financeiro. A Figura 7 mostra um exemplo desses gráficos juntos em uma só visualização. A exibição desses dois gráficos concomitantemente facilita a visualização e a correlação entre as duas análises (variação do valor e volume

de negociação). Isso permite que o investidor identifique se os resultados do gráfico de volume de negociação tem relação com os resultados do gráfico de variação do valor.



Figura 7 – Gráfico de Velas Com Volume

Para exibir os resultados desse trabalho, optou-se pela construção de dois gráficos principais, sendo o primeiro um gráfico de vela (exibindo a variação do valor do *Bitcoin*, e um segundo gráfico de barras indicando o sentimentos sobre a criptomoeda. Além desses dois gráficos principais, a plataforma irá exibir outros dois gráficos secundários, realizando uma comparação entre o número de *tweets* classificados como positivos e negativos em relação ao preço de fechamento da criptomoeda.

Um exemplo da exibição dos gráficos principais pode ser visualizado na figura 8.



Figura 8 – Gráfico de Velas Vs Sentimento

Na construção dos gráficos, foi utilizada a biblioteca **Plotly**. Essa biblioteca é uma ferramenta de construção de gráficos e oferece suporte para os mais diversos tipos de gráficos. Além disso, essa ferramenta foi escolhida por facilitar a publicação e hospedagem dos gráficos online. Após a construção e *deploy* (implantação) do gráfico, é possível gerar um iframe ou trecho html e disponibilizar seu gráfico em qualquer aplicação web.

A atualização do gráfico é feita por uma Cron. Uma Cron consiste um recurso do sistema operacional que permite agendar tarefas e executar programas em momentos específicos. Por meio deste recurso, é possível agendar a atualização do gráfico, deste modo, a cada período de hora, um novo gráfico é construído e enviado para **Plotly**.

3.5.4 Interface Gráfica

Nas figuras 9 e 10, pode-se visualizar as duas páginas principais da ferramenta. A página *Home*, que possui uma descrição do projeto, incluindo informações acerca de ferramentas utilizadas, informações de contato e links para o repositório do projeto, e a página de gráficos, contendo os gráficos de vela e de sentimento. Para melhorar o design da aplicação, foi utilizado o *framework* de estilo **materialize**.

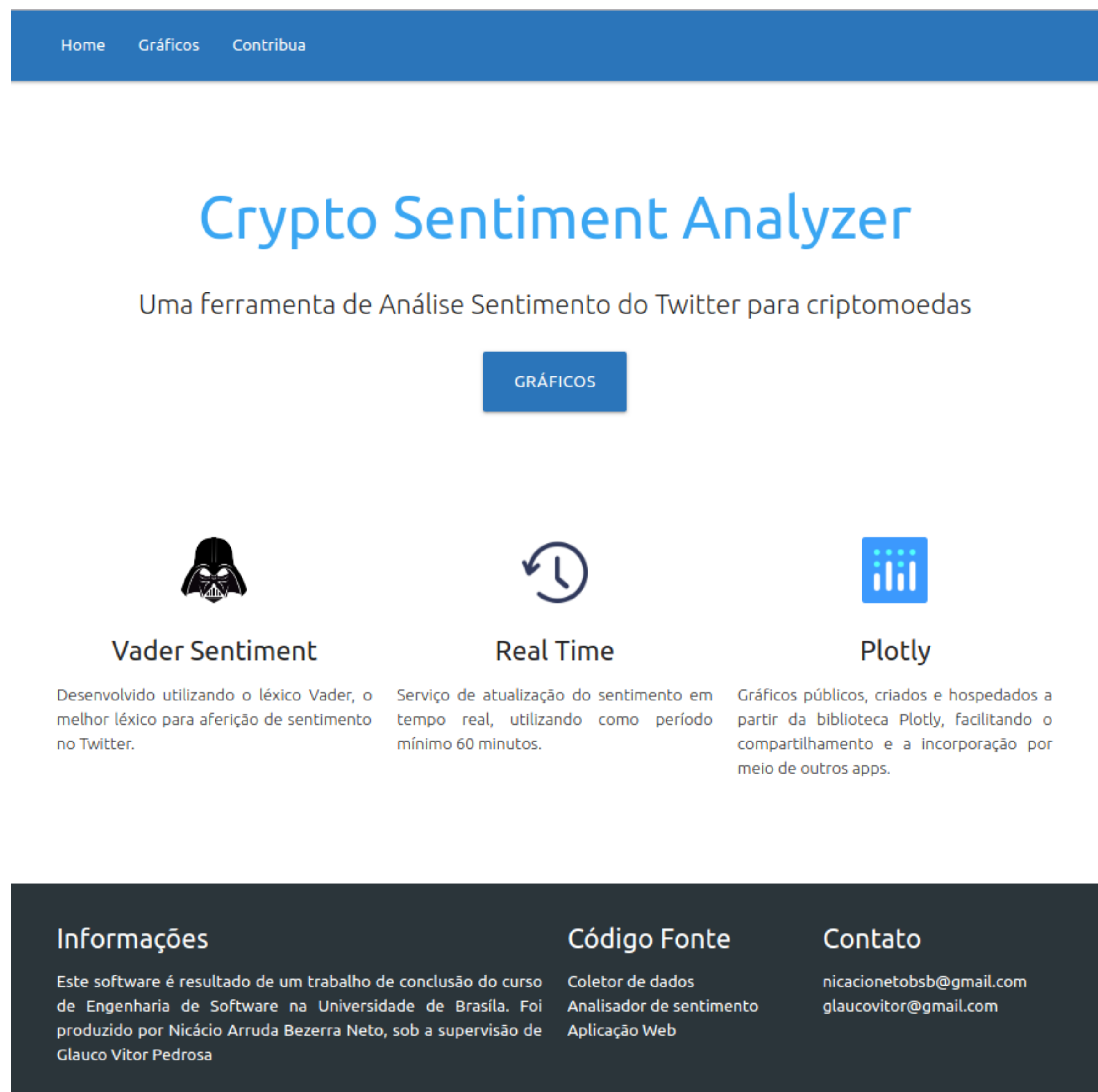


Figura 9 – Página Home

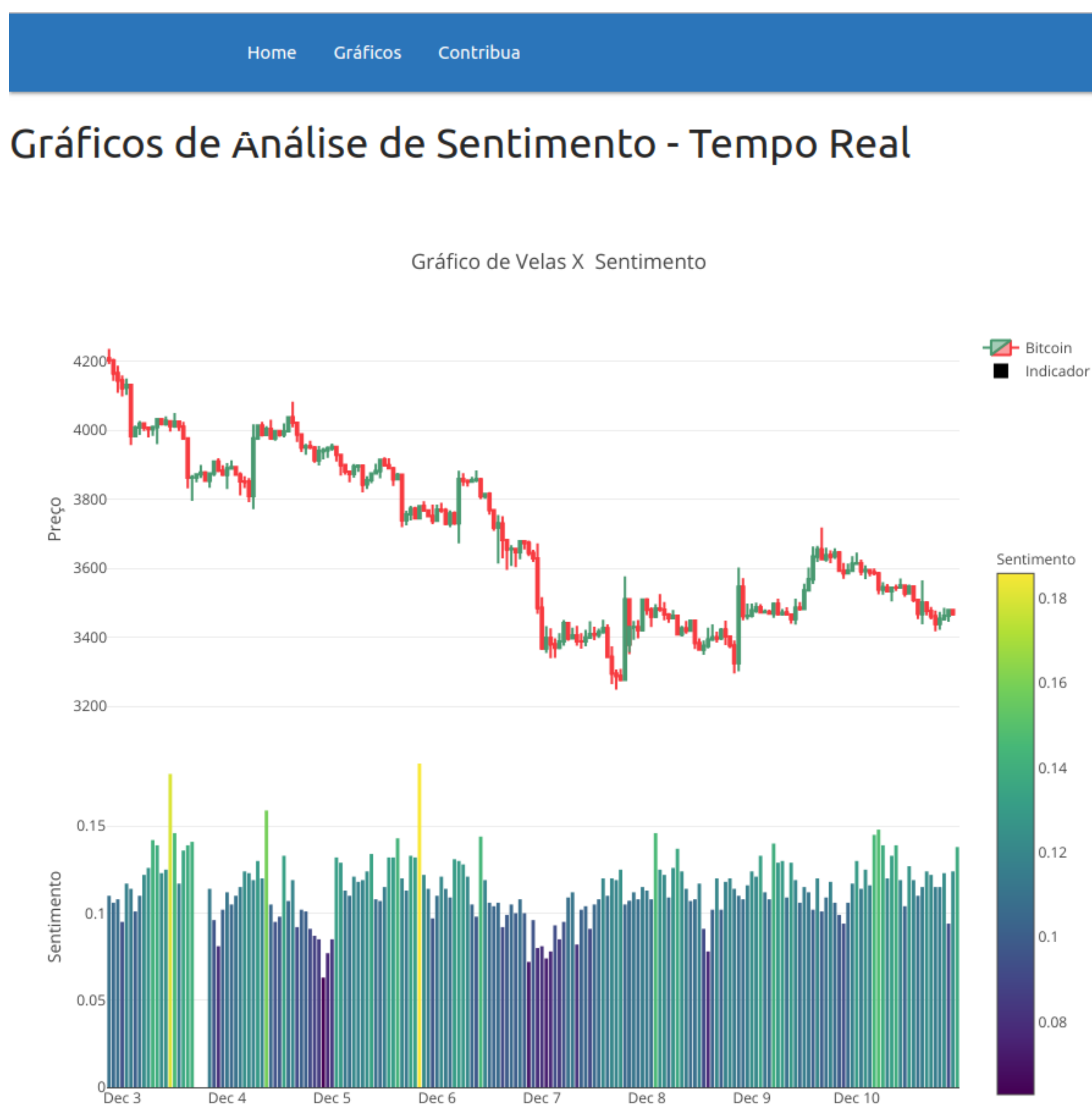


Figura 10 – Página de Gráficos

3.5.5 Ferramentas

A aplicação Web foi desenvolvida utilizando a linguagem **Ruby** na versão 2.5.3 juntamente com o *framework* **Rails** na versão 5.2.1.

No ambiente de produção, foi utilizado o servidor **Passenger** como servidor Web e o servidor **Nginx** como proxy reverso.

O servidor **Nginx** é responsável por receber diretamente as requisições HTTP e encaminhar para o servidor **Passenger**, que irá interagir diretamente com a aplicação Web.

O servidor **Nginx** foi escolhido como servidor de proxy reverso por conta de sua performance. O servidor **Phusion Passenger** foi escolhido como servidor Web por conta de ser o servidor mais utilizado em aplicações desenvolvidas com o frameword **Rails**, e por ser de fácil integração com o **Nginx**. No frontend, foi utilizado o *framework* CSS **materialize**.

3.6 Arquitetura da Solução

A Figura 11 mostra um diagrama esquemático representando a arquitetura da solução desenvolvida.

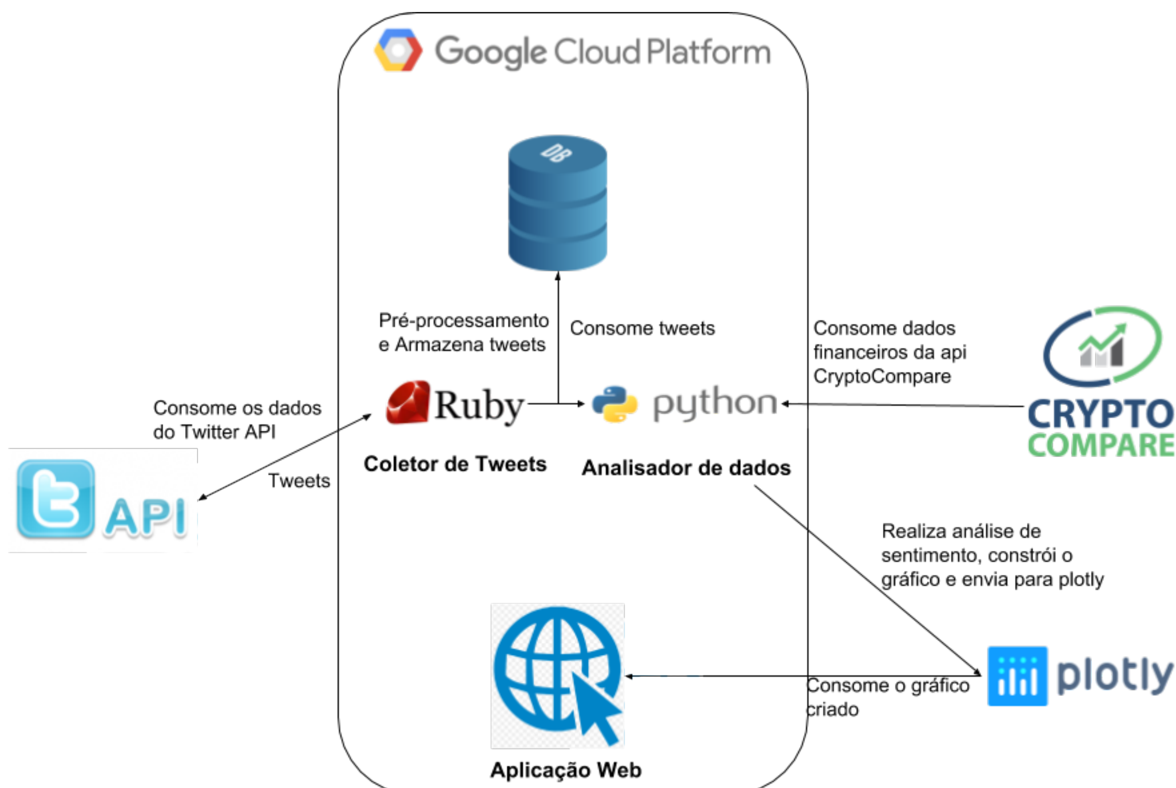


Figura 11 – Esquemático do projeto

Nesta arquitetura, nota-se que os dados oriundos da *API* do Twitter são recebidos pelo Coletor de Tweets. Este software é o responsável por se comunicar com *API*, recebendo as publicações, além de realizar o pré-processamento e armazenamento dos dados. O Analisador de dados ordena esse conjunto de publicações por data. Após a ordenação, ele busca as informações financeiras relacionadas ao *Bitcoin* consumindo a *API* da **API Cryptocompare**. Com essas informações ele cria os gráficos de vela e de barras e realiza o *deploy* (Implantação) no serviço de hospedagem de gráficos **Plotly**.

Durante todo esse processo, a aplicação web está consumindo o gráfico do serviço **Plotly**. O Coletor de Tweets e o Analisador de dados, bem como a aplicação web estão alocados em um servidor dedicado, hospedado no serviço de nuvem *googlecloud*.

Todo o código produzido encontra-se no repositório público de projetos GitHub:

- Coletor de Tweets
- Analisador de dados
- Aplicação Web

A aplicação desenvolvida pode ser acessada pelo seguinte endereço: <http://104.196.46.178:3000/>.

4 Resultados

O objetivo dessa seção é mostrar algumas análises realizadas através dos gráficos gerados pelo sistema desenvolvido e discutido na seção anterior. A ideia é identificar correlações entre a variação do valor do *Bitcoin* e a quantidade de sentimentos (positivos e/ou negativos) sobre a criptomoeda na plataforma Twitter.

4.1 Análise 1

A Figura 12, mostra o gráficos comparando o valor de fechamento do *Bictoin* e a quantidade de *tweets* negativos sobre a moeda no período entre 01:00hr do dia 21/10/2018 até às 0:00hrs do dia 31/10/2018. Nota-se pelo gráfico que o valor do *Bitcoin* permaneceu com valor estável entre os dias 21/10/2018 e 29/10/2018, porém, as 10:00hrs do dia 29/10/2018, sofreu uma variação negativa de 2,61% no seu valor (de $6459 para $6290). Neste mesmo período, houve um aumento no número de *tweets* negativos de 479 para 753, representando um aumento de 57.20%.

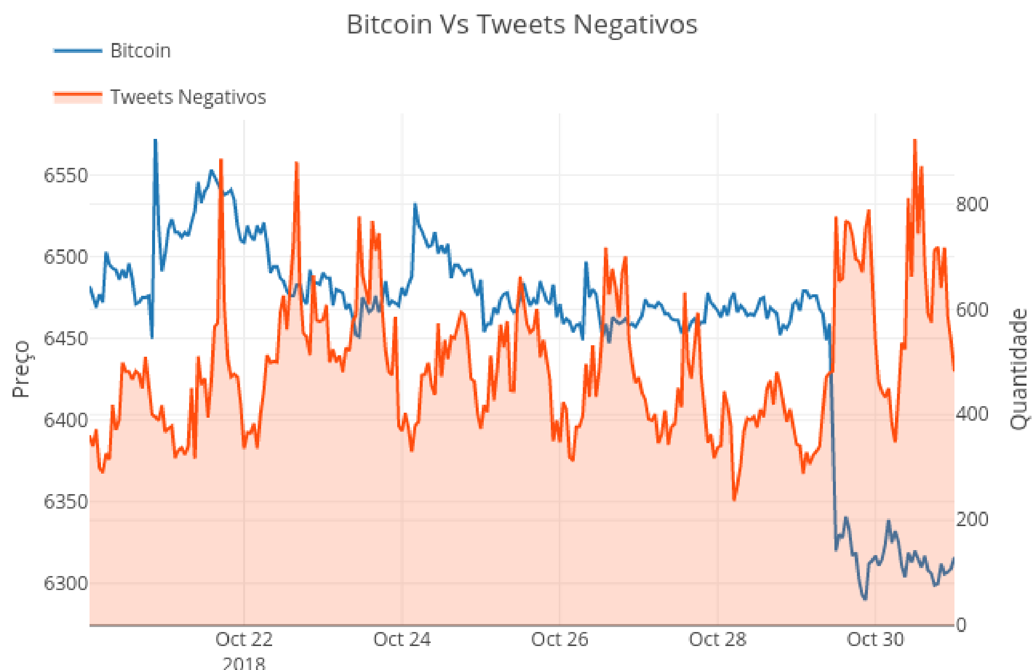


Figura 12 – Sentimento Negativo 21/10/2018 - 31/10/2018

Foi possível verificar, através dessa análise, que o crescimento da quantidade de *tweets* negativos acompanhou a queda do valor do *Bitcoin*. Não foi possível identificar

com bastante antecedência que o valor do *Bitcoin* iria cair, porém nota-se uma tendência de estabilidade no valor da criptomoeda quando há uma redução na quantidade de *tweets* negativos. A partir dessa observação, foi realizada uma nova análise que será discutida a seguir considerando também a quantidade de *tweets* positivos.

4.2 Análise 2

Sob a perspectiva de outro período (01/09/2018 01:00hrs à 16/09/2018 0hrs), as figuras 13 e 14 mostram a quantidade de *tweets* positivos e negativos da criptomoeda, em comparação com o valor de fechamento do *Bitcoin*.

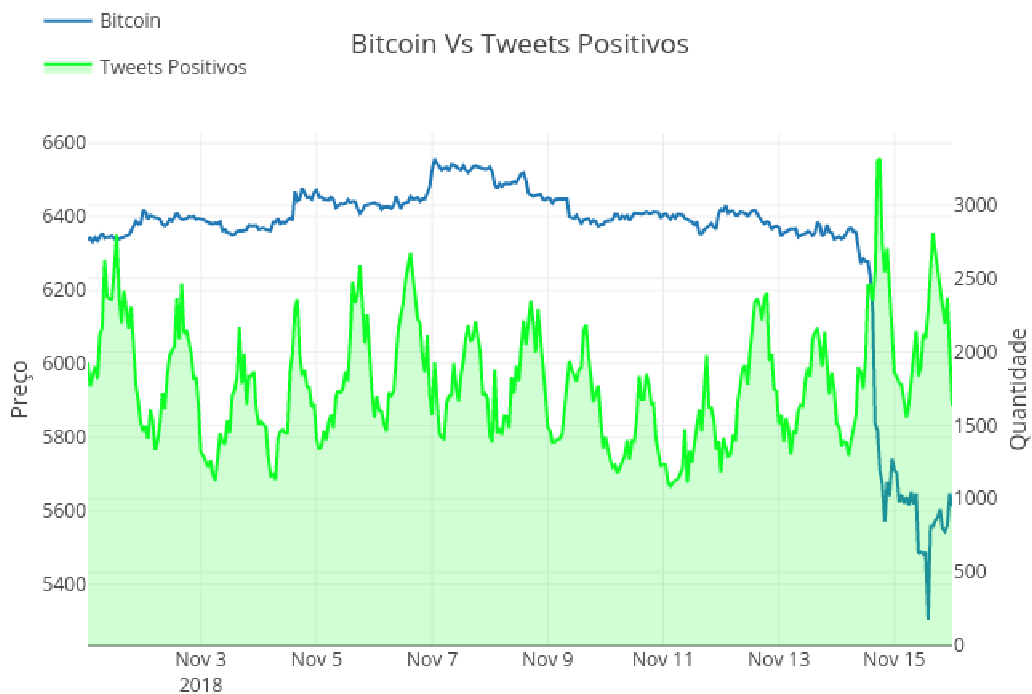


Figura 13 – Sentimento Positivo 01/11/2018 - 16/11/2018

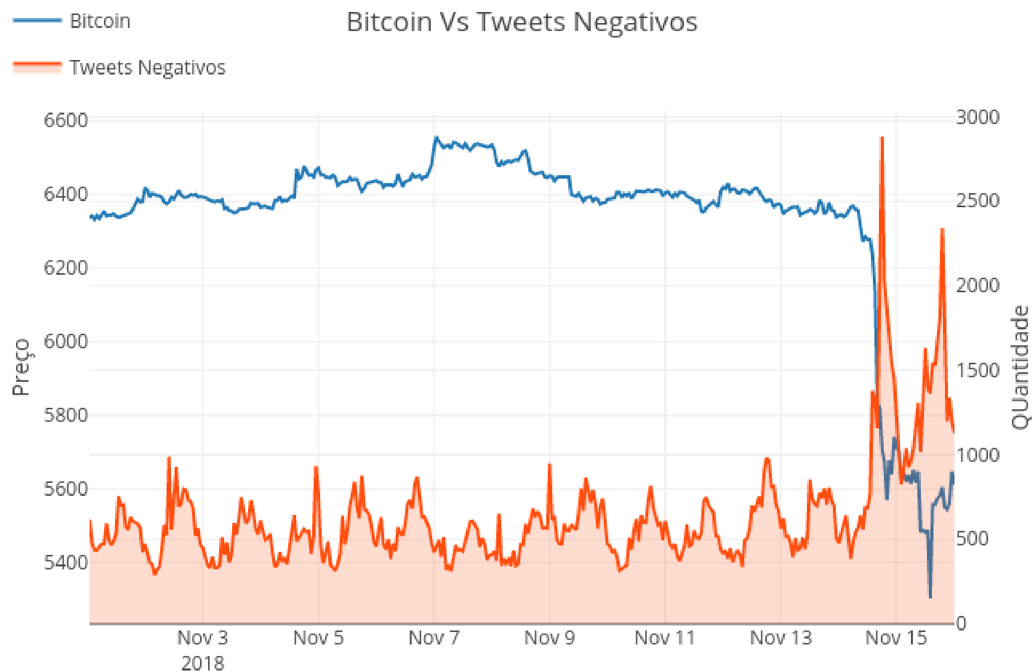


Figura 14 – Sentimento Negativo 01/11/2018 - 16/11/2018

Para este período, nota-se o mesmo comportamento que o período da Análise 1. O *Bitcoin* estava com valor estável até o dia 14/11/2018 às 08:00hrs, porém iniciou uma queda constante de $6358 à $5570 às 20:00hrs. Essa redução foi de 12.39%, enquanto o valor de *tweets* negativos cresceu consideravelmente de 570 para 1829, representando 220.88%. Porém, desta vez, foi possível notar um aumento no número de *tweets* positivos de 1559 para 2544, representando 63.18%. É perceptível que, mesmo após ás 20:00hrs, horário em que o preço ficou relativamente estável, o número de *tweets* positivos e negativos continuou a subir, chegando a 3314 e 2881 respectivamente.

Notou-se que a análise a partir do número de publicações poderia ser diretamente influenciada pelo horário, visto que existem horários com mais pessoas *online* e, por isso, mais publicações são realizadas. Para termos uma análise confirmatória sobre esse fato, foi utilizado o indicador definido na seção 3.4.3, pois ele realiza o balanceamento do sentimento como uma única medida que é dividida pelo número de *tweets* publicados. Lembrando que, quanto menor o *indicador* mais negativo é o sentimento. E esta nova análise, considerando esse *indicador* será discutido a seguir.

4.3 Análise 3

Nas figuras 15 e 16, é realizada uma comparação entre o valor de fechamento da moeda e o indicador de sentimento definido na seção 3.4.3.

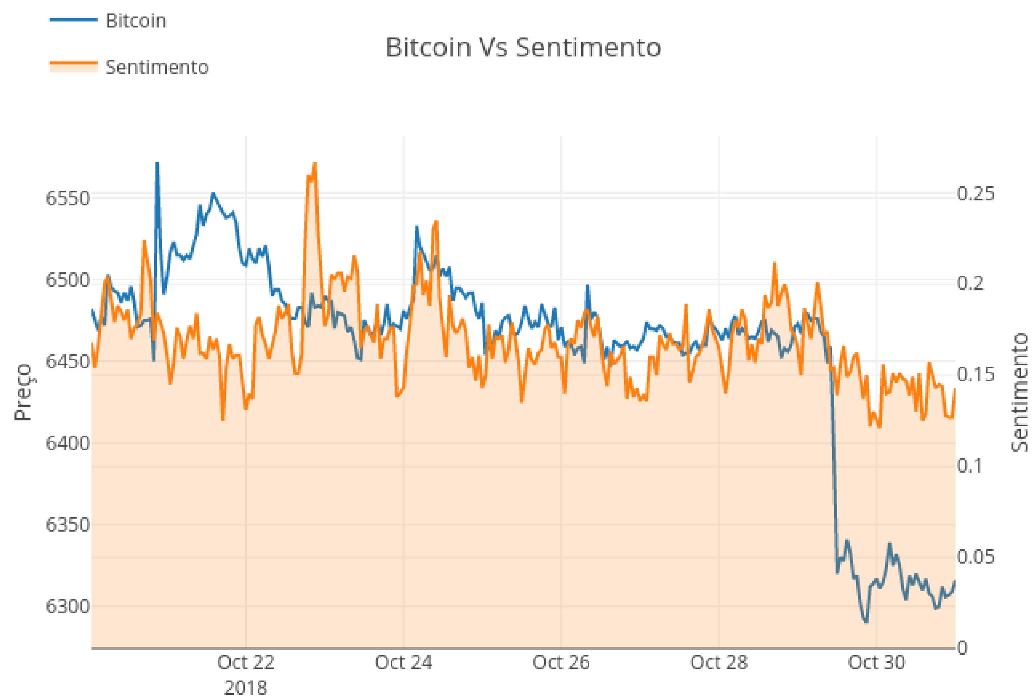


Figura 15 – Indicador 21/10/2018 - 31/10/2018

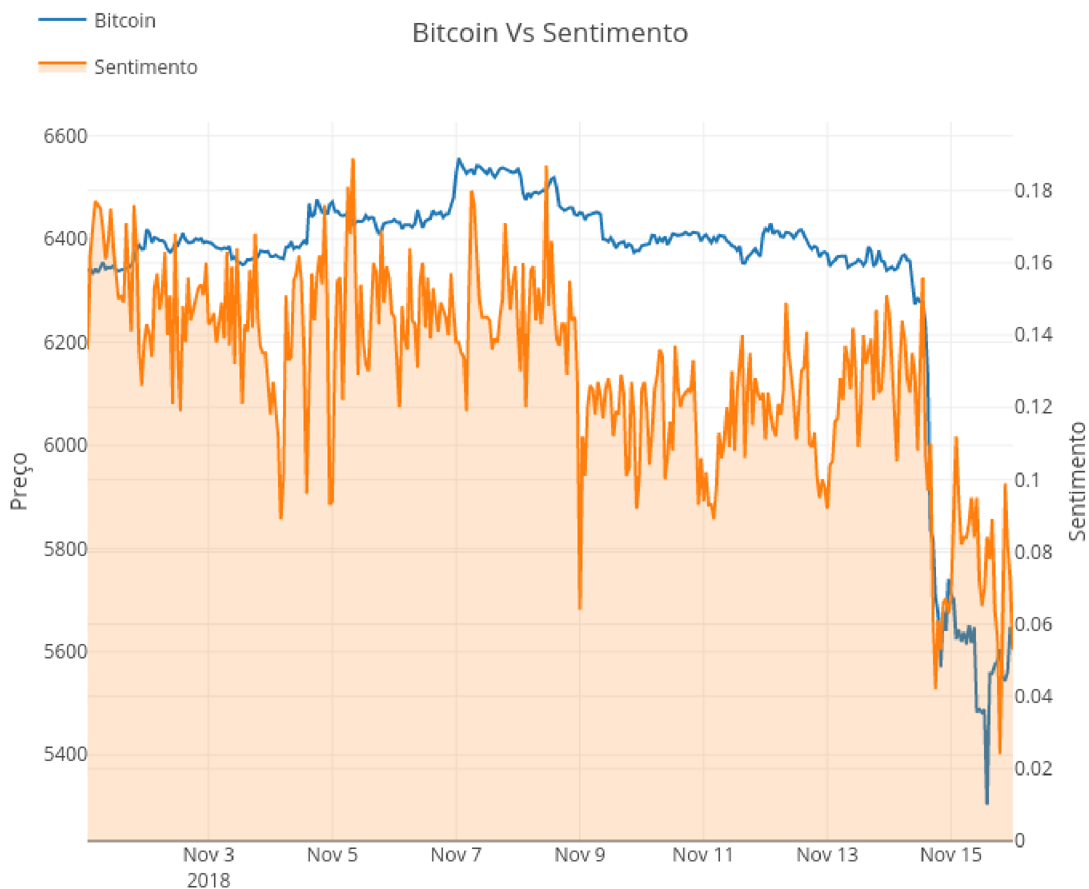


Figura 16 – Indicador 01/11/2018 - 16/11/2018

A partir da análise das figuras 15 e 16, percebe-se que o sentimento negativo aumentou a medida em que o preço da moeda diminuía, repetindo o mesmo comportamento dos gráficos anteriores.

Através dessa análise preliminar, observou-se que a mudança da opinião pública a respeito da moeda sofreu variação após a variação do valor da moeda. Além disso, o número de publicações aumenta como um todo após grandes variações. **Essas constatações fortalecem o pensamento de que existe correlação entre o preço e o sentimento.**

Todavia, somente com a análise realizada não foi possível identificar uma estratégia **preditiva**, de modo que, o sentimento medido refletiu as variações que já haviam ocorrido, ou que estavam ocorrendo no presente. Entretanto, é importante salientar que, o aumento do número de *tweets* negativos mesmo que não identificado com antecedência, pode ser usado como indicador de venda, visto que, esse aumento mostrou uma tendência de queda do valor da moeda.

As análises realizadas até o momento se basearam em horas e talvez esse período não fosse o suficiente para que o *Bitcoin* incorporasse o valor do sentimento público em seu preço. Por isso, foi realizada uma nova análise considerando outro intervalo de tempo, que será discutida a seguir.

4.4 Análise 4

Com base nos resultados das análises anteriores, decidiu-se utilizar um período diário. Utilizando um período de 23 dias, a Figura 17 exibe o gráfico do valor de fechamento em comparação ao indicador de sentimento.

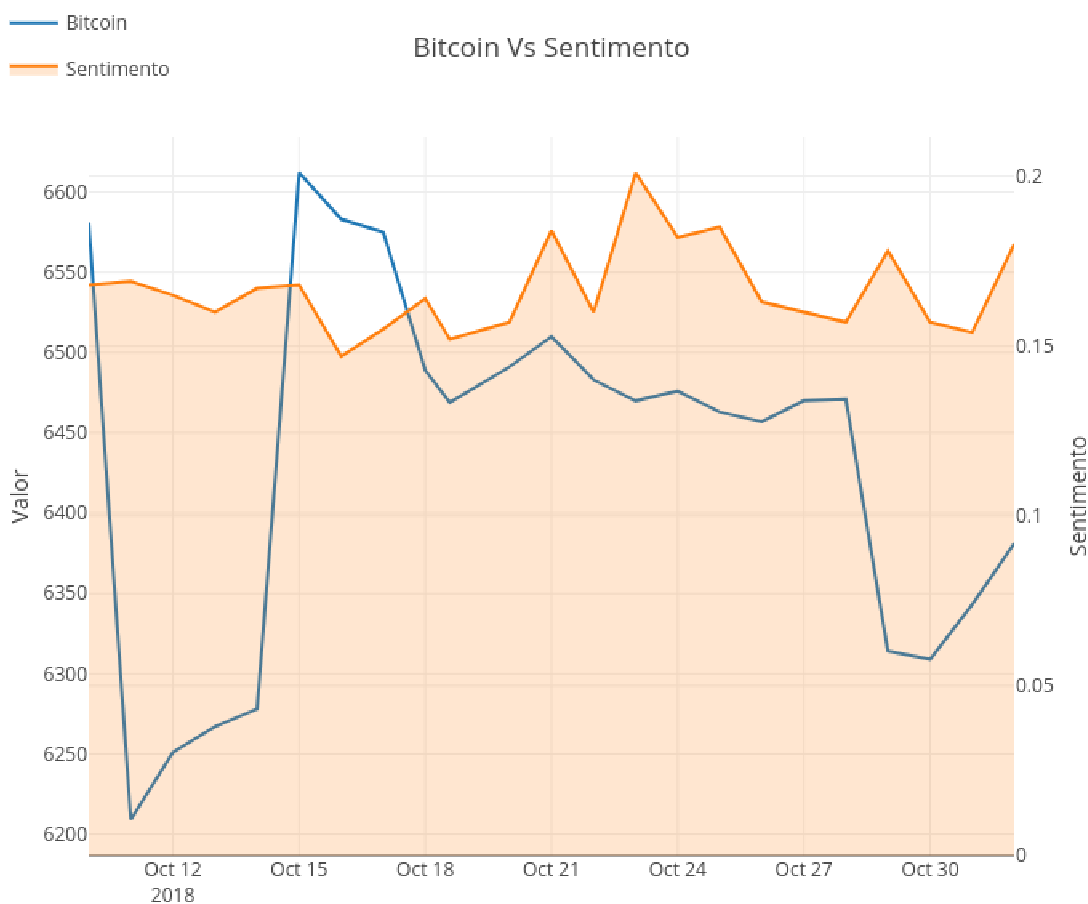


Figura 17 – Sentimento 09/10/2018 - 31/10/2018

A partir da visualização do gráfico, presente na figura 17, percebe-se que a moeda se recuperou de uma queda brusca, começando $6272 no dia 14 para $6612 no dia 15, enquanto que o indicador só começou a crescer no dia 19, atingindo a máxima no dia 22. Este resultado reforçou a ideia de que, a análise do sentimento no *Twitter* reflete o

comportamento passado ou presente da moeda, de modo que, o sentimento acompanhou o aumento do valor da moeda com certo atraso.

5 Conclusão

5.1 Considerações Finais

Na literatura encontra-se trabalhos relacionando o comportamento dos usuários nas redes sociais e a movimentação no mercado do *Bitcoin*. Motivado por essa constatação, este trabalho se propôs a desenvolver uma ferramenta computacional capaz de realizar, de forma automática, a análise de sentimentos sobre a cripto-moeda *Bitcoin* usando o *Twitter*. A ideia foi construir uma plataforma que pudesse ser utilizada por investidores como suporte preditivo para a volatilidade do *Bitcoin*.

Ao fim deste trabalho, realizou-se com sucesso a análise de sentimento dos usuários do *Twitter* em relação ao *Bitcoin*, coletando e filtrando as publicações relacionadas ao tema e utilizando o léxico *Vader*, aliado a técnicas de processamento de linguagem natural para aferir o sentimento presente nas publicações.

Pode-se observar que existe, de fato, uma relação entre o preço e o sentimento expresso pelos usuários, porém, este sentimento representa o comportamento passado, ou presente da moeda, tendo em vista que, não foi possível identificar com bastante antecedência uma mudança de sentimento que ocasiona-se uma mudança no preço da moeda. Desse modo, não foi possível comprovar que, de fato, a variação do preço é causada pelo sentimento, reforçando o entendimento que, o sentimento expresso pelos usuários consiste, em grande parte, de uma reflexão do mercado passado/atual.

Por fim, conseguiu-se construir um software para tornar a tarefa de coleta e análise automatizada. possibilitando assim, a disponibilização dessas informações na Web. Apesar de que, as informações disponibilizadas não possam ser utilizadas unicamente para propósito de previsão, pode-se observar que o sentimento acompanhou os movimentos nos preços da moeda, indicando os anseios, medos e desejos dos usuários e investidores, sobretudo, podem indicar também, movimentos de alta e de baixa nos preços.

5.2 Trabalhos Futuros

A realização desse trabalho trouxe consigo grandes aprendizados, sobretudo, em mineração de dados, análise de sentimento e processamento de linguagem natural. Os conhecimentos adquiridos serão úteis para os trabalhos futuros que poderão decorrer deste. Um destes trabalhos é a evolução da ferramenta de coleta e análise de *tweets*, permitindo com que essa ferramenta possa trabalhar com outras moedas e ativos financeiros.

Na plataforma Web, algumas melhorias podem ser feitas, uma delas é a criação

de filtros por períodos para melhoria da visualização dos gráficos, além de disponibilizar rotas de API, para que pesquisadores e pessoas interessadas possam consumir os dados utilizados para gerar os gráficos.

Referências

BEGUM, H.; BEGUM, S. J. Data mining tools and trends – an overview. In: . [S.l.: s.n.], 2013. ISSN 2278-9359. Citado na página 25.

BUKOVINA, J.; MARTICEK, M. *Sentiment and Bitcoin Volatility*. [S.l.], 2016. Disponível em: <https://ideas.repec.org/p/men/wpaper/58_2016.html>. Citado 2 vezes nas páginas 19 e 32.

CAGINALP, G.; LAURENT, H. The predictive power of price patterns. *Applied Mathematical Finance*, Routledge, v. 5, n. 3-4, p. 181–205, 1998. Disponível em: <https://doi.org/10.1080/135048698334637>. Citado na página 30.

COOLEY, R.; MOBASHER, B.; SRIVASTAVA, J. Web mining: Information and pattern discovery on the world wide web. In: . [s.n.], 1997. p. 558–567. Cited By 632. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031385045&partnerID=40&md5=3594959d057c1d82599b7389f7b4534e>. Citado na página 25.

GARCÍA, D.; SCHWEITZER, F. Social signals and algorithmic trading of bitcoin. *CoRR*, abs/1506.01513, 2015. Disponível em: <http://arxiv.org/abs/1506.01513>. Citado na página 31.

GARCIA, D. et al. The digital traces of bubbles: feedback cycles between socio-economic signals in the bitcoin economy. *Journal of The Royal Society Interface*, The Royal Society, v. 11, n. 99, 2014. ISSN 1742-5689. Disponível em: <http://rsif.royalsocietypublishing.org/content/11/99/20140623>. Citado na página 31.

GEORGOULA, I. et al. Using time-series and sentiment analysis to detect the determinants of bitcoin prices. 2015. Citado na página 31.

HEARST, M. A. Untangling text data mining. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 1999. (ACL '99), p. 3–10. ISBN 1-55860-609-3. Disponível em: <https://doi.org/10.3115/1034678.1034679>. Citado na página 25.

HUTTO, C. J.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *ICWSM*. [S.l.: s.n.], 2014. Citado 2 vezes nas páginas 27 e 38.

KAMINSKI, J.; GLOOR, P. A. Nowcasting the bitcoin market with twitter signals. *CoRR*, abs/1406.7577, 2014. Disponível em: <http://arxiv.org/abs/1406.7577>. Citado na página 31.

KARALEVICIUS, V.; DEGRANDE, N.; WEERDT, J. D. Using sentiment analysis to predict interday bitcoin price movements. *The Journal of Risk Finance*, v. 19, n. 1, p. 56–75, 2018. Disponível em: <https://doi.org/10.1108/JRF-06-2017-0092>. Citado na página 32.

KARPPI, T.; CRAWFORD, K. Social media, financial algorithms and the hack crash. *Theory, Culture & Society*, v. 33, n. 1, p. 73–92, 2016. Disponível em: <https://doi.org/10.1177/0263276415583139>. Citado 2 vezes nas páginas 19 e 23.

KRISTOUFEK, L. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. In: *Scientific reports*. [S.l.: s.n.], 2013. Citado 2 vezes nas páginas 19 e 30.

KWAK, H. et al. What is twitter, a social network or a news media? In: . [s.n.], 2010. p. 591–600. Cited By 3131. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77954619566&doi=10.1145%2f1772690.1772751&partnerID=40&md5=42cc3dac9d8543fcf8691d3707051e0c>. Citado na página 23.

LIU, B. *Sentiment Analysis and Opinion Mining*. [S.l.]: Morgan & Claypool Publishers, 2012. ISBN 1608458849, 9781608458844. Citado na página 26.

MARSHALL, B. R.; YOUNG, M. R.; CAHAN, R. Are candlestick technical trading strategies profitable in the japanese equity market? *Review of Quantitative Finance and Accounting*, v. 31, n. 2, p. 191–207, Aug 2008. ISSN 1573-7179. Disponível em: <https://doi.org/10.1007/s11156-007-0068-1>. Citado na página 30.

MATTA, M.; LUNESU, M. I.; MARCHESI, M. Bitcoin spread prediction using social and web search media. In: *UMAP Workshops*. [S.l.: s.n.], 2015. Citado na página 31.

NISON, S. *Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Techniques of the Far East*. New York Institute of Finance, 1991. ISBN 9780139316500. Disponível em: <https://books.google.com.br/books?id=nz57QgAACAAJ>. Citado na página 28.

RANCO, G. et al. The effects of twitter sentiment on stock price returns. *PLOS ONE*, Public Library of Science, v. 10, n. 9, p. 1–21, 09 2015. Disponível em: <https://doi.org/10.1371/journal.pone.0138441>. Citado na página 19.

ZAFARANI, R.; ABBASI, M. A.; LIU, H. *Social Media Mining: An Introduction*. New York, NY, USA: Cambridge University Press, 2014. ISBN 1107018854, 9781107018853. Citado na página 25.